RMetS
Royal Meteorological Society

Development and validation of observing-system simulation experiments at NASA's Global Modeling and Assimilation Office

Ronald M. Errico,[a,b]* Runhua Yang,[c] Nikki C. Privé,[a,b] King-Sheng Tai,[d] Ricardo Todling,[b] Meta E. Sienkiewicz[d] and Jing Guo[d]

[a]*Goddard Earth Sciences Technology and Research Center, Morgan State University, Baltimore, MD, USA*
[b]*Global Modeling and Assimilation Office, National Aeronautics and Space Administration, Greenbelt, MD, USA*
[c]*IM Systems Group, Inc., Rockville, MD, USA*
[d]*Science Systems and Applications, Inc., Greenbelt, MD, USA*

*Correspondence to: R. M. Errico, Global Modeling and Assimilation Office, NASA/GSFC Code 610.1, Greenbelt, MD 20771, USA. E-mail: ronald.m.errico@nasa.gov

Initial design and validation of baseline Observing System Simulation Experiments (OSSEs) at NASA's Global Modeling and Assimilation Office (GMAO) are described. The OSSEs mimic the procedures used to analyze global observations for specifying states of the atmosphere. As simulations, however, OSSEs are not only confined to already existing observations and they provide a perfect description of the true state being analyzed. These two properties of the simulations can be exploited to improve both existing and envisioned observing systems and the algorithms to analyze them. Preliminary to any applications, however, the OSSE framework must be adequately validated.

This first version of the simulated observations is drawn from a 13 month simulation of nature produced by the European Center for Medium-Range Weather Forecasts. These observations include simulated errors of both instruments and representativeness. Since the statistics of analysis and forecast errors are partially determined by these observational errors, their appropriate modelling can be crucial for validating the realism of the OSSE. That validation is performed by comparing the statistics of the results of assimilating these simulated observations for one summer month compared with the corresponding statistics obtained from assimilating real observations during the same time of year. The assimilation system is the three-dimensional variational analysis (GSI) scheme used at both the National Centers for Environmental Prediction and GMAO. Here, only statistics concerning observation innovations or analysis increments within the troposphere are considered for the validation.

In terms of the examined statistics, the OSSE is validated remarkably well, even with some simplifications currently employed. In order to obtain this degree of success, it was necessary to employ horizontally correlated observation errors for both atmospheric motion vectors and some satellite observed radiances. The simulated observations with added observation errors appear suitable for some initial OSSE applications. **Copyright © 2012 Royal Meteorological Society**

Key Words: OSSE; data assimilation; atmospheric observations

Received 7 March 2012; Revised 9 July 2012; Accepted 24 July 2012; Published online in Wiley Online Library

Citation: Errico RM, Yang R, Privé NC, Tai K-S, Todling R, Sienkiewicz ME, Guo J. 2012. Development and validation of observing-system simulation experiments at NASA's Global Modeling and Assimilation Office. *Q. J. R. Meteorol. Soc.* DOI:10.1002/qj.2027

1. Introduction

An observing system simulation experiment (OSSE) is a numerical experiment conducted with a data-assimilation system (DAS) and numerical prediction model that traditionally uses simulated rather than real observations. These are drawn from some dataset representing the states to be observed. For an OSSE applied to the atmosphere, this is most appropriately a temporal sequence of atmospheric fields generated by a sufficiently realistic simulation model, termed a 'nature run' (NR). The simulated observations are then ingested by the DAS. Various metrics are applied to quantify the accuracy of the analyses produced, particularly standard ones measuring fits to observations and forecast skill. The impacts of various configurations of the observing system can then be compared. Unlike an observing system experiment (OSE) conducted with real observations, an OSSE is not limited to using observations that currently exist.

OSSEs are most often employed to estimate quantitatively the potential improvements in climate analysis and weather prediction to be gained by augmenting the present atmospheric observing system with additional envisioned types of observations that do not yet exist. In particular, the utilities of competing designs of proposed observing systems can be compared within the context of modern data-assimilation systems. OSSEs can also be used to explore some otherwise elusive characteristics of an already existing observing system such as, for example, its analysis error statistics as in Errico *et al.* (2007a). Essentially, these latter goals are achievable because in the OSSE framework a dataset representing the atmosphere or ocean is precisely known, unlike the case of the real atmosphere, thereby allowing an explicit and precise determination of analysis errors.

Before conducting an OSSE to investigate proposed observing systems, there are several reasons to conduct baseline experiments simulating a current observing system. Most importantly, as a simulation, any particular OSSE framework should first be validated by comparing corresponding metrics of the DAS and subsequent forecasts applied to equivalent real observations (termed 'GDAS' here) in order to establish its credibility. Most proposed observing systems are intended to be augmentations of the existing (baseline) systems that they are supposed to improve, so a baseline validation is especially relevant. Many past OSSEs have been criticized privately because their validations have been insufficient or even absent, sometimes resulting in their gross misinterpretation. The complexity of modern data-assimilation systems, cost of modern observing systems and importance of accurate state estimation require that OSSEs be carefully validated, based explicitly on comparisons of metrics applied to corresponding results in real and simulated frameworks.

The GMAO OSSE development began as an offshoot of what has been called the 'Joint OSSE', which was an informal collaboration of several investigators in Europe and the United States (Masutani *et al.*, 2007). In particular, the European Center for Medium-Range Weather Forecasts (ECMWF) created a usefully realistic NR dataset for the atmospheric OSSE community to use. Since then the GMAO effort has evolved to become independent of the joint OSSE, aside from use of this NR.

We begin with a description of the NR and our DAS in the next section. This is followed by descriptions of the ways in which observations and their associated errors are simulated and validated. A sample of validation results is presented in section 6. These are all in terms of background and analysis fits and differences, with forecast metrics discussed in a companion paper (Privé *et al.*, 2012). A summary and plans for further development are then offered.

2. The nature run and DAS

The ECMWF NR representing 'truth' for the OSSE is provided by a 13 month forecast for the period 10 May 2005–31 May 2006. This uses the ECMWF operational model from 2005 (version cy31r1) with resolution defined by triangular spectral truncation at wavenumber 511 with 91 levels in the vertical above the surface ($T511L91$). It is a 'free-running' solution, in the sense that the only information provided about the real atmosphere, besides the model equations, parametrization and initial conditions, are the sea-surface temperature (SST) and sea-ice fraction provided as lower boundary conditions. The latter are derived from an analysis of real SST and sea-ice observations during that period.

The NR datasets are provided on a reduced linear Gaussian grid with fields defined at 512 latitudes and between 1024 and 18 longitudes, with the smaller numbers closer to the poles. The vertical coordinate is a hybrid one: terrain-following sigma near the surface, pressure above 150 hPa and a blending of the two in between. The simulated fields are available every 3 h for both prognostic and diagnostic model fields. In several aspects, this NR appears adequately realistic (Reale *et al.*, 2007), but some shortcomings are also apparent (e.g. too few high-level clouds, as shown in McCarty *et al.*, 2012).

The DAS used here is an upgraded version of that described by Rienecker *et al.* (2008). It includes the Grid-point Statistical Interpolation (GSI) scheme for three-dimensional variational analysis originally developed at NCEP (Wu *et al.*, 2002; Kleist *et al.*, 2009) and now jointly developed with the GMAO (version 5.7.1). The incremental analysis update of Bloom *et al.* (1996) is used for enhanced dynamical balance. Both real observation and OSSE assimilation experiments are run with horizontal resolution 0.5° latitude by 0.625° longitude. Atmospheric fields are defined on 72 levels above the surface using the same hybrid coordinate (denoted as η) as in the NR, but with a different blending of pressure and sigma levels specified. All results shown in terms of fields will be on η-surfaces denoted in terms of what the corresponding pressure surface would be if the surface pressure were 1000 hPa everywhere. Radiative transfer calculations in GSI are performed using release 1.2 of the Community Radiative Transfer Model (CRTM; Han *et al.*, 2006).

When the DAS is applied to either simulated or real observations, it uses identically specified background-error and observation-error statistics previously tuned for an earlier version of the system applied to real data. The same quality control and data-selection algorithms are also used. The validation here thereby determines how well the OSSE can match a real analysis using as similar a DAS as possible.

The assimilation model is version 5 of the Goddard Earth Observing System model (GEOS–5) as used at the GMAO (but not at NCEP; Rienecker *et al.*, 2008). It uses a finite-volume time-splitting algorithm for its dynamics (Lin, 2004), in contrast to the semi-Lagrangian, semi-implicit

formulation in the ECMWF model. The parametrized physics (Bacmeister *et al.*, 2006) also differs from that at either NCEP or ECMWF.

The OSSE validation is performed for the period 1–31 July 2005. The OSSE actually begins at 0000 UTC on 15 June 2005, however, to allow the spin-up of radiance bias correction coefficients (Dee, 2005) and background errors. The initial background fields are the result of a 2 day assimilation that itself begins with a real background for 0000 UTC on 13 June 2005 but has as its observations simulated rawinsondes drawn from every second latitude, longitude and vertical level of the NR every 6 h, with no observation errors added. These high-density, high-accuracy observations are used to accelerate adjustment of the background toward the NR state. Although the subsequent 17-day spin-up period is short with respect to the 1 month *e*-folding time employed for computing the radiance bias-correction coefficients, this is deemed sufficient because, unlike for real observations, the OSSE simulated ones have no sources of large bias and all the coefficients are initialized to spun-up values from previous experiments that were only slightly less well tuned.

3. Simulation of observations

Observations are simulated by applying an appropriate algorithm (an observation operator) to the NR fields at instants in time. While it may seem attractive to begin by simulating current observations in the most realistic ways possible, this is not a generally prudent approach for several reasons: (1) the fidelity of some realistically simulated observations may be detrimentally affected by unrealism in some aspect of the NR (e.g. as will occur if radiances are computed using a good model for scattering by clouds but the high-level NR clouds are deficient); (2) the most realistic observation operators available may still be too physically deficient (e.g. although a radiance-scattering model may generally describe cloud effects well, it may treat poorly optically thin clouds that affect those radiance observations actually retained by quality-control (QC) algorithms); (3) information on a data type or spatial scale required by a realistic observation operator may be absent from the NR (e.g. the cloud microphysical properties required by a radiation-scattering model); (4) the DAS may not consider effectively some aspects of the observations that characterize their realism, and thus great effort may be expended in simulating complex aspects for which a much simpler treatment may be adequate (as in the examples provided in what follows); (5) some aspects of realism are expected to have little impact on the time- and space-averaged metrics to be provided either for this validation or for other studies planned for the near-term, and thus have little importance at this stage of development. Our approach has been to start by using simple observation operators, adding more realism as required to improve the validation of the OSSE framework.

When creating simulated observations for an OSSE, it is useful to distinguish between the modelling of observational signal and noise. The signal provides information that is both interpretable and desirable to be retained by the DAS, whereas any noise will be a source of error that the DAS will tend to diminish. The latter includes gross errors, which often, but not always, lead to data rejection by QC algorithms, and also typically smaller errors of instruments and representativeness that are less easily detected. Different DAS may alter the partitioning between signal and noise; e.g. when a DAS is changed from computing cloud-free radiances using only assumed cloud-free observations to actually computing the effects of clouds in order also to retrieve cloud properties. The latter requires attention to details of cloud effects, whereas the former essentially considers clouds only as a source of error. It would naturally be best if the observations were indeed simulated as realistically as possible so that the products would be useful for both present and future applications, independent of what is presently considered as the signal. This approach would require much greater expertise and attention to detail than we can now offer, however, and would not be relevant in the DAS available to us. That admission guides our present distinction between signal and noise, with the latter primarily treated as suitably tuned random errors to be added.

3.1. *Simulation of conventional observations*

'Conventional' observations include all those not provided to the GSI as measures of radiances. Specifically, these include measurements by rawinsondes, dropsondes, pilot balloons, aircraft (AIREP, PIREP, ASDAR and MDCRS), wind profilers, Doppler radar (VAD) winds, ships, land stations (for surface pressure only) and ocean buoys. They also include winds estimated by tracking features in geostationary satellite images of cloud or water vapour (denoted collectively as 'SATWND') and by SSM/I and QuikScat sea-surface winds (denoted collectively as 'SSWND' here). The operational version of GSI also uses precipitation retrieved from satellite observations, but since these have negligible impact on the metrics used for validation here and insufficient validation of instantaneous fields of NR precipitation has been conducted, their use in the OSSE has been deferred. The only *in situ* surface measurements considered by GSI are radiosonde temperatures and surface pressures, meso-net surface pressures and measurements by ships or buoys.

The location and time for each simulated conventional observation are specified as those of the corresponding real observation that was considered by the operational GMAO DAS for that same date. The real data have been partially thinned and quality-controlled and thus are devoid of many, but not all, of the gross observational errors that generally occur. By assigning locations and times of simulated data in this way, several aspects of the spatial and temporal distribution of these data are constrained to be identical to those of real observations, with some important exceptions described later.

The only observation operators used for the considered conventional observations are spatial and temporal interpolations from the NR gridded fields to the observation locations and times. Interpolations are bilinear in horizontal distance, log-linear in vertical pressure and linear in time. Surface observations are produced at the interpolated NR topographic height or 10 m above that, depending on the observation type.

Three aspects of conventional observations that may be important but not critical are neglected. One is that the locations of 'significant-level' radiosonde observations are specified by the locations of the corresponding real data rather than by applying the rules for reporting such observations to the NR fields. Thus, data for significant-level

reporting that would otherwise occur given the NR fields may be absent, and vice versa. A second aspect is that locations for wind observations determined by tracking features (e.g. clouds) detected in real satellite imagery may be inconsistent with the locations of such features in the NR. Thus, cloud-track wind observations may be simulated where the NR has no clouds. Third, the SSWND observations are simply determined by horizontally and temporally interpolating from the 10 m winds provided in the NR data set, without reconsideration of momentum profiles in the vertical.

There are also some neglected aspects that are considered unimportant. Locations of soundings (e.g. by rawinsondes) are treated as being fixed to their initial locations (i.e. without accounting for their drift). Observations of precipitation rates are neglected because they have negligible impact on the metrics presented here and the quality of their usefulness is questionable even in the GDAS (Errico *et al.*, 2007b).

3.2. Simulation of radiance observations

The baseline dataset includes brightness temperatures (T_b) simulated for the HIRS/2 (on NOAA–14), HIRS/3 (on NOAA–16,17) and AIRS (on AQUA) infrared (IR) sensors and for the AMSU–A (on NOAA–15,16 and AQUA), AMSU–B (on NOAA–15,16,17) and MSU (NOAA–14) microwave (MW) sensors. GOES radiances, which have been shown currently to have negligible impact in the DAS, are not used in this GDAS or OSSE. At all observed locations, vertical profiles of relevant fields are determined by horizontal and temporal interpolation of fields on the NR data levels. These are then passed to the CRTM. The consequences of using a very similar radiative transfer model to both simulate and assimilate the observations are described in the next section.

The GSI thins the set of available radiance observation locations for computational reasons concerning both the CRTM and the conditioning of the DAS minimization algorithm. This is accomplished by first dividing the area of the globe into a set of approximately equal-area trapezoids with lengths of sides ranging from 145–240 km, depending on the instrument. For each trapezoid, instrument and satellite considered, it selects observations for just one location from the set of all observed locations within that trapezoid. The selection is based on an estimate of what may be 'best' in terms of the location, time and least undesirable effects of clouds. For AIRS observations, the selection algorithm also favours locations that are associated with nearby AMSU–A observations in the same data report. The CRTM is then used to simulate observations only at these locations. The simulated radiance observations for the OSSE are pre-thinned similarly to the way in which GSI subsequently does so, but using trapezoids with sides approximately 45 km, so that the GSI then still makes its own data selection among the up to 25 simulated observations provided to it.

Both the data selection and data QC algorithms in GSI act to discard radiance observations suspected of being adversely cloud-affected, since its CRTM algorithm treats all radiative transfer as cloud-free. Many radiance observations for channels with weighting functions that peak sufficiently higher than the altitudes of presumed clouds, however, are retained. The primary effect of these procedures is therefore to reduce the numbers of assimilated observations, especially in cloudy regions for lower-peaking channels. A minimum requirement for the OSSE is to incorporate a similar cloud effect so that realistic counts and spatial distributions of observations are obtained for each instrument channel.

For all the reasons stated at the beginning of section 3, at the current stage of development, this primary cloud effect on IR radiances has been introduced in a simple tunable manner. First, the existence of a significant cloud effect at each observation location is determined by a simple probabilistic function of the fractional areas of high, medium and low-level cloud covers provided in the NR dataset. These determinations employ tunable functions that allow for the possibility of 'holes' in clouds and for adjustments due to possible unrealism of the instantaneous fractional cloud distributions in the NR. If a significant cloud is specified to be at an observation location, it is treated as a black body and the radiative transfer calculations are performed with the cloud top as the radiatively effective surface without the additional consideration of scattering by clouds below. At most locations, this will result in cold brightness temperatures for the affected channels, which will be recognized by the QC as undesirable cloud effects. The probabilistic function is crudely tuned to yield monthly-mean counts of the GSI quality-accepted observations that are similar to those obtained for real observations for each IR channel. The same function, albeit with different parameters, is used for all instruments. Further details are provided in Appendix A.

A more subtle effect of clouds occurs when they are optically thin with respect to a radiance calculation; i.e. when they affect the radiative transfer but not in such a gross way as to be detected by the QC. In this case, the difference between the cloud-free CRTM calculation and the real cloud-scattering result is interpreted by the GSI as an error of representativeness. Such an error of representativeness is also introduced using the OSSE simulation procedure, but we do not claim that important characteristics of the real and OSSE errors are similar. We do compensate for some discrepancy, however, in the way additional simulated observation errors are introduced in the OSSE.

Effects of precipitation on MW transmission are treated analogously to cloud effects on IR transmission. In this way, the GSI QC excludes consideration of MW observations in regions of sufficiently strong precipitation. Details appear in Appendix A.

For IR and especially MW radiances, the treatment of surface emissivities in the created observations presents a special problem. Emissivity modelling for real observations has many uncertainties, exacerbated by poorly described or unknown surface inhomogeneities (e.g. regarding soil and vegetation properties). The CRTM in both the OSSE simulations and GSI is provided with a vegetation-type index (10 types) and vegetation fraction, snow depth and temperature, soil moisture content and skin temperature. The latter three quantities are provided by horizontally interpolated values from the NR. The vegetation values used for the OSSE observation simulation are taken as the nearest-gridpoint values provided on an approximately half-degree Cartesian grid and are similar but not identical to those values provided in the GSI. Over water, the 10 m winds interpolated from that NR field are used rather than recalculated winds derived from stability profiles. The modelled emissivity used to simulate the observations is therefore not identical to that used in the GSI assimilation, but the difference is likely much less than

the emissivity-responsible representativeness error implicit when assimilating real observations. This discrepancy can be ameliorated by adding a simulated representativeness error to the OSSE observations, but only if the character of the real error is sufficiently known. There is likely a spatially varying but temporally correlated component to the real error (e.g. due to seasonal variations of subgrid vegetation) for which no error-modelling attempt has been made for the OSSE. Presently MW channels strongly affected by the surface are not used, so the lack of sufficient emissivity error in the OSSE framework is somewhat moot here.

No attempt has been made to simulate error biases in satellite observations explicitly. The biases for which we know the characteristics are the same biases that the GSI is designed to effectively remove. Since other biases are presumably smaller but not sufficiently known, at this stage of development there is little reason to simulate any biases. The OSSE retains a radiance-bias correction in the DAS, however, since the CRTM versions for simulating and assimilating are not identical. As when assimilating real observations, some of the innovation bias will actually be due to background error bias but incorrectly attributed to observation error bias and corrected as such. This radiance-bias correction therefore also serves as a source of some error that can occur in both OSSE and GDAS contexts.

4. Simulation of observation errors

The statistics of analysis errors are partly determined by the statistics of observation errors, as revealed by the fundamental DAS equations and as examined further by Daley and Menard (1993). The observation errors can be considered as sums of two distinct contributions: instrument errors and representativeness errors. The first is absent from the simulated observations because no physical instruments are involved. The other, although partially present, is likely diminutive for reasons offered below. If the OSSE is to be validated by comparing measures of its behaviour with corresponding results when assimilating real data, then it is imperative to employ it with the same prescribed background and observation-error covariances (respectively **B** and **R** in the standard notation of Ide *et al.*, 1997) and to ensure that the simulated observations have errors with statistics, especially covariances, similar to real ones. Only under these conditions can a meaningful validation be performed.

The representativeness errors have various sources. One is the DAS's limited representation of the true atmospheric state in terms of a finite set of grid points or spatial functions. The other consists of errors in formulations of the physics relating what is observed to what is analyzed, most notably errors in determining satellite-observed radiances in terms of profiles of temperature, moisture, ozone, etc. Both sources may be considered as arising from errors in the forward models, including spatial interpolation schemes and radiative transfer schemes applied in the DAS. Some of this error is implicitly present within the observations simulated for the OSSE, since the NR and DAS represent the atmosphere differently. They also differ in the values of some parameters used for radiative transfer. The spatial resolutions of the NR and DAS, however, are not extraordinarily different. Neither are the versions of the CRTM applied to the NR and within the DAS. Variances of the representativeness errors introduced implicitly in this way should therefore be significantly smaller than the corresponding variances present when real observations are assimilated.

For the above reasons, errors are added to the observations simulated from the NR. For each data type, these are determined from random distributions that may be horizontally, vertically or channel-correlated or uncorrelated. The variances for these distributions depend only on data type and pressure or channel (i.e. not geographical location). They are determined by a tuning procedure so that corresponding temporal variances of observation innovations $\mathbf{d} = \mathbf{y}^{\mathrm{o}} - H(\mathbf{x}_{\mathrm{b}})$, where $\mathbf{y}^{\mathrm{o}}$ are observation values and H is a corresponding operator that yields an estimate of $\mathbf{y}$ from the background-state estimate $\mathbf{x}_{\mathrm{b}}$, are matched in the OSSE and GDAS contexts. This tuning assumes that any mismatch of $\mathbf{d}$ statistics is due to inappropriate observation error. This is likely the dominant cause for reasons previously presented, but misapplication due to mismatched error statistics of $\mathbf{x}_{\mathrm{b}}$ cannot be discounted. After the $\mathbf{d}$ statistics are tuned, other metrics must be investigated to check the reasonableness of the assumptions made about observation errors. The tuning procedure is described in Appendix B and its success is indicated in section 6. For the distributions of all explicitly added errors, the means are zero, so no additional biases are introduced.

We reasonably assume that the additional observational errors in conventional, closely-spaced, significant-level sounding data should be weakly correlated. For simplicity, we use a Gaussian-shaped correlation function defined as

$$\rho(p_1, p_2) = \exp\left[-0.5\left\{\frac{R_{\mathrm{g}} T_{\mathrm{o}}(\ln p_1 - \ln p_2)}{g L_{\mathrm{v}}}\right\}^2\right],$$

where ρ is the correlation, p_1 and p_2 are any two pressure levels, R_{g} here is the gas constant, $T_{\mathrm{o}} = 270\,\mathrm{K}$ is an approximate tropospheric mean temperature, g here is the acceleration of gravity and L_{v} is a vertical decorrelation length-scale (500 m for wind and temperature and 180 m for relative humidity).

For all AMSU, HIRS and MSU instruments, the added errors have both uncorrelated and horizontally correlated components. The horizontal structures of the correlations are Gaussian, each prescribed by a length-scale L. The fraction of total error variance contributed by the latter is denoted ν. Both L and ν are tuned analogously to the procedure for observation-error variances separately for each instrument, satellite and channel. So, for these instruments, no error correlations between channels are assumed. From these parameters, global random correlated fields are produced from which realizations of the errors are horizontally interpolated. The random fields are generated by projection of spherical harmonic functions defined using random (complex) spectral coefficients. The desired spatial Gaussianity is assured by constraining the expected power spectrum of the latter to be given by the shape appearing in Appendix C. The random coefficients themselves are drawn from normal distributions of mean 0 and appropriate variances.

For AIRS, being a hyper-spectral instrument, the explicitly added observational errors are correlated between channels. This is accomplished by creating horizontally random fields as for the other radiance observation errors but instead using these to define random coefficients for the principal

components of the channel-error covariance matrix. The latter is estimated from examination of observation increment covariances obtained from an assimilation of real data. The channel correlations resemble those appearing in Bormann *et al.* (2010). No horizontal correlations are considered because they appear small and are otherwise difficult to tune in conjunction with the channel correlations.

Analogous to the treatment of AIRS, the added errors of SATWNDS utilize principal components of an assumed vertical covariance of observation errors. The vertical correlations are defined as in (1) with $L_v = 750$ m. The simultaneously vertically and horizontally correlated fields are defined on a grid with high spatial resolution so that when its random values are interpolated to observation locations substantial variance is not lost due to averaging of random numbers (e.g. if values for locations centred between two points where the fields are randomly defined with no correlation are determined by linear interpolation, the resulting variance at those locations will be half that of the random field defined on the grid). In the present OSSE system, errors for the two wind components are determined separately, using the same horizontal Gaussian-shaped correlations as for radiances. The L and ν are determined by tuning similar to that for radiances. Both wind components use the same error parameters.

4.1. *Tuned standard deviations for added observation errors*

Examples of tuned standard deviations of the observation errors explicitly added in the OSSE are presented in Figure 1 for four observation types. Also presented for each are the square roots of corresponding elements of **R** specified in the GSI. The values reported for the OSSE do not include the implicit portions of representativeness errors that are due to the differing NR and GDAS resolutions, etc.

The standard deviations for AMSU–A on NOAA–15-added OSSE errors (Figure 1(a)) are approximately one half of the GDAS error-table values. Similar ratios are obtained for other satellite instruments, except for AIRS for which one-third is a more typical ratio. For rawinsonde values of T_v (Figure 1(b)), the ratios are close to one, with minima of 0.79 at 700 and 250 hPa and maxima of 1.4 in the stratosphere. Ratios of rawinsonde values for wind (Figure 1(c)) are also close to 1, with a maximum of 1.34 near the surface and minimum of 0.93 for $p < 50$ hPa. For SATWND (Figure 1(d)), ratios range from 0.32 near the surface to 0.48 for $p < 200$ hPa. For most conventional observations not shown, the assigned ratio is close to 1, except for surface marine and buoy observations, for which ratios are near 0.5.

4.2. *Tuned horizontal correlations for added observation errors*

Examples of the tuning parameters for L and ν for the horizontally correlated components of the explicitly added observation errors appear in Figure 2. For AMSU–A on NOAA–15, values are presented only for the channels actually used by the GSI. The SATWND values are those tuned for GOES IR and visible winds, but are applied to SATWND observations deduced from other geostationary satellite images as well.

For AMSU–A, L (Figure 2(a)) increases from 320 to 610 km as the channel number increases (channel ten being a slight exception). This increase also corresponds to an increase of the height at which each channel's weighting function peaks. For channels 5–7, $\nu < 0.07$ (Figure 2(b)), indicating that negligible correlated error is added for these channels. For other channels, however, $0.23 < g < 0.41$, with the largest value occurring for channel 4, which is most affected by the Earth's surface.

For SATWND, L (Figure 2(c)) is largest (450 km) near the surface and a minimum (220 km) near 400 hPa. Values of ν (Figure 2(d)) range from 0.24 at $p = 850$ hPa to 0.69 for $p < 200$ hPa. For $p < 750$ hPa, most of the added error for SATWND is therefore horizontally correlated.

5. Validation metrics

The OSSE has been validated using many metrics. Only such metrics as can also be obtained from the GDAS are examined here. So, for example, scores of how well the OSSE analyses compare with the (known) NR truth are deferred to a future report because there is no corresponding truth for the GDAS. Also, attention is focused on the troposphere rather than on the surface, stratosphere or mesosphere. Other metrics or portions of the atmosphere will be addressed in later reports.

For this baseline set of OSSE observations, parameters used in the observation-error models have been selected in an attempt to match particularly standard deviations of both innovations and analysis increments in the OSSE with corresponding ones in the GDAS. It is produced using the same period, resolution and DAS, drawing from the same set of real observations employed to determine observation locations in the OSSE.

The part of the cost function that measures the misfit between observations and implied background values is denoted as

$$J_o = \mathbf{d}^T \mathbf{R}^{-1} \mathbf{d} \,,$$

where superscript 'T' denotes a transpose. In the GSI, **R** is diagonal, so J_o is simply a sum of squares of the innovations for each individual observation weighted by the corresponding inverses of the assumed variances of the observation errors.

Statistics of analysis increments are also examined. These are determined as

$$\mathbf{x}_a - \mathbf{x}_b = \mathbf{K}\mathbf{d},$$

where $\mathbf{x}_a$ is the analysis and **K** is the GSI approximation to the Kalman gain matrix resulting from its iterative algorithm. Thus, the analysis increments are weighted sums of the innovations.

6. Results

All these results are evaluated from four times daily analysis over the 31 days of July. Statistics are computed using global averages unless specified otherwise.

6.1. *Numbers of observations assimilated*

By design, the numbers of observations of various types actually assimilated within the OSSE after quality control are very similar to the corresponding numbers in the GDAS. Table 1 presents these numbers for indicated observation types as averages per 6 h assimilation period for the month

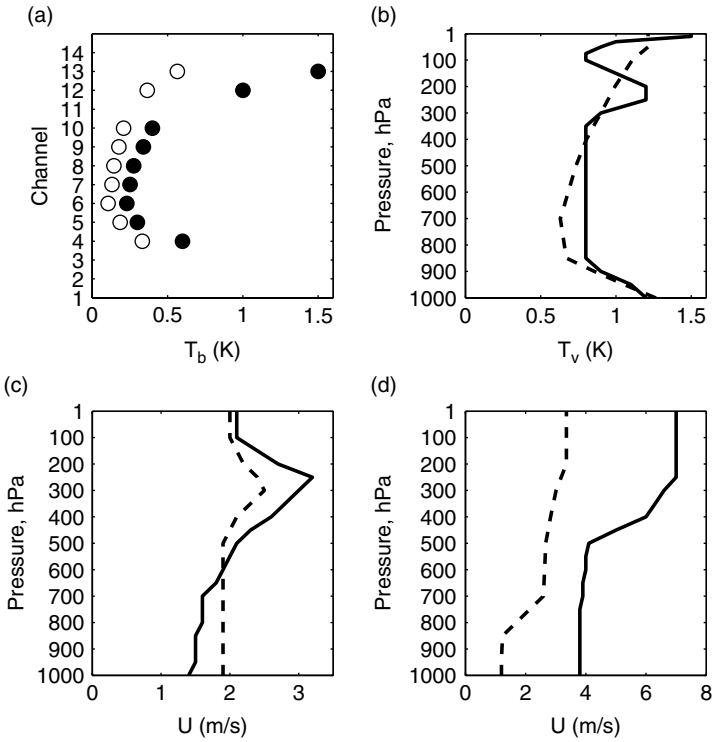

Figure 1. Comparison of standard deviations of the simulated observation errors explicitly added in the OSSE (dashed lines and open circles) with the corresponding values (solid lines and filled circles) specified in the GSI. Values are shown for (a) T_b for AMSU–A on NOAA–15, (b) T_v for rawinsondes, (c) wind components for rawinsondes and (d) SATWND.

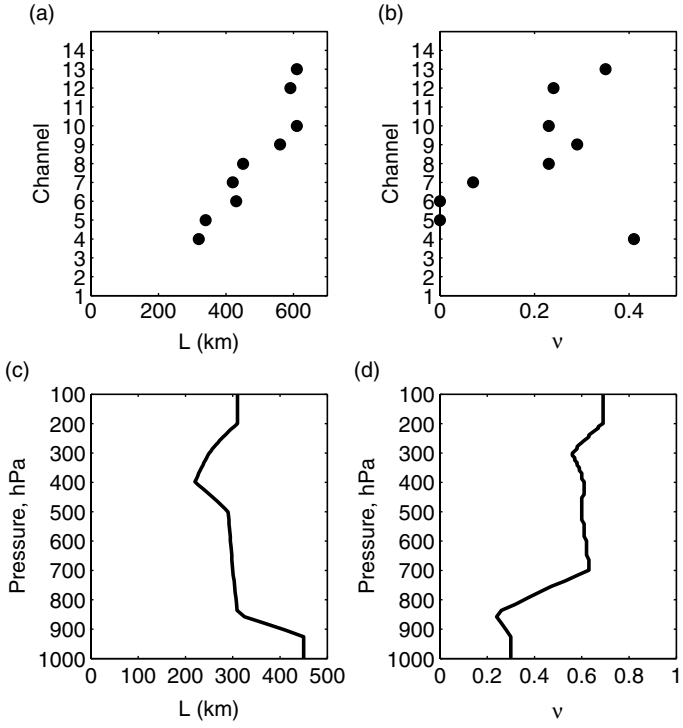

Figure 2. Tuned values for (a,c) the length-scales of horizontally correlated components of the explicitly added observation errors in the OSSE and (b,d) the fractions of the total added error variance contributed by the horizontally correlated component. Values are for (a,b) AMSU–A on NOAA–15 and (c,d) SATWND.

of July. The conventional observation category 'Sondes' includes rawinsondes, pibals and dropsondes.

For all conventional observation types, the corresponding numbers differ by less than 4% because (1) a simulated observation has been produced for every real observation that has passed a preliminary quality-control check and (2) no non-Gaussian gross errors have been added to the simulated values. The corresponding numbers are not necessarily identical because additional quality control is performed during the assimilation procedure.

Unlike for the conventional observations, the OSSE radiance observation counts depend on the algorithm used

to introduce cloud and precipitation effects. As one set of metrics of that algorithm's performance, the mean numbers of accepted observations per day as a function of channel for AIRS appear in Figure 3. They are all within 10% of each other, which is rather remarkable given the crude way cloud effects are incorporated in the OSSE radiance observations. For other satellite instruments, the degrees of correspondence are similar.

A typical example of the geographical distribution of simulated observations after data thinning and removal of those identified by the DAS quality control as likely cloud-affected is shown in the top panel of Figure 4. Specifically, this is for channel 295 on AIRS–AQUA (corresponding to channel 106 in the NCEP subset designation), for the 6 h period centred on 1800 UTC, 12 July 2005. This channel is potentially affected by clouds at any level, which explains why observations are absent within patches that otherwise are clearly within the observation swaths. The spatial density of the observations outside these cloudy patches is reflective of the satellite data thinning incorporated within GSI.

The result corresponding to the top panel of Figure 4 but for real observations appears in the bottom panel. Observation locations in the two panels are not in one-to-one correspondence. Even in cloud-free regions, the data thinning will not choose identical observation locations from the two datasets. Even if temporal and spatial means of cloud distributions for the NR are entirely realistic, the locations of clouds at any specific time may differ (since the OSSE and GDAS weather differ), as is apparent in Figure 4. This difference is only slightly compounded by use of a random function to determine where OSSE observations see through holes in clouds. The proper way to compare the panels in Figure 4 is therefore to consider their qualitative characteristics. This includes the typical observation densities as well as the sizes of patches of rejected observations. Ideally, this comparison should be conducted by placing the OSSE result among, for example, ten GDAS results for the same month, and then attempting to discern which is the OSSE result. Given this figure, that task should be sufficiently difficult. The number of assimilated observations shown in Figure 4 for GDAS is 3% greater than for the OSSE.

6.2. *Contributions to cost function*

Contributions to J_o for a 6 h assimilation period averaged for July for various types of observations appear in Table 1 for both OSSE and GDAS results. For all radiance types, corresponding values are within 5% of each other. For conventional observations the agreement is typically less, with the discrepancies mostly in the range 15–20% of the GDAS values. A single exceptionally poor agreement is for the SSWND, where the OSSE value is more than three times the GDAS one and where further examination reveals specifically that the QuikScat winds are the main discrepancy. If the contributions to the sum of J_o excluded these winds, the agreement between OSSE and GDAS values would be within 2%.

6.3. *Standard deviations of innovations*

Values of J_o are convenient for comparing innovations aggregated over observations with typical magnitude values that vary greatly, since that metric normalizes each contribution by its observation-error variance. In this subsection, standard deviations of innovations are presented for aggregates over more homogeneous data subsets.

Table 1. Mean numbers of observations assimilated and contributions to J_o per 6 h period during July.

Obs. type	Number GDAS	Number OSSE	J_o GDAS	J_o OSSE
Radiances				
MSU	13464	14081	3563	3625
AMSU–A	231413	226903	34563	33546
AMSU–B	15556	15380	8783	8377
HIRS/2–3	68887	69171	17410	18999
AIRS	415746	394567	162474	156717
Conv. *T*				
Sondes	10923	10872	20219	17546
Aircraft	30646	30394	32024	44877
Sfc. Marine	2387	2392	683	1090
Conv. wind				
Sondes	28872	28559	48001	39763
Aircraft	69091	68634	88535	102156
SATWND	42900	42958	13153	14590
Sfc. Marine	4820	4844	3511	4936
SSWND	90636	90642	29468	102096
Prof./VAD	47677	48490	15306	11593
Conv. *q*				
Sondes	5553	5443	3996	4168
Sfc. Marine	1011	1012	210	250
Conv. p_s				
Land	42168	42691	7301	6944
Ocean	4336	4167	1072	1138
Sum	1126086	1101200	495160	576373

Standard deviations for innovations of virtual temperature (T_v), specific humidity (q) and eastward wind component (u) for rawinsonde observations north of 20°N, appear in Figure 5 as functions of pressure for OSSE and GDAS assimilations. Note that for almost all tropospheric levels and fields, OSSE values are within a few per cent of their GDAS counterparts. The only exceptions are for T_v at 925 hPa and for q below 600 hPa, where the discrepancies are as large as 25%. The latter subset is not yet part of the automated observation-error tuning process.

The standard deviations for v innovations of GOES IR and visible winds (derived from tracking cloud features) evaluated north of 20°N also appear in Figure 5. This figure shows discrepancies of less than 10%, except at 925 hPa for which the sample size is small. Corresponding results for u indicate almost no differences because the observation-error parameters were tuned using the u discrepancies from a previous experiment.

Statistics for other conventional observations (not shown) display less agreement than those for rawinsondes and cloud-tracked winds. This is mostly because less effort was employed to tune their added observation-error statistics, since they have less impact within the DAS than other observations. The OSSE values for these types are generally within 10% of their GDAS counterparts. The two exceptions are SSM/I ocean-surface wind speeds, for which the OSSE

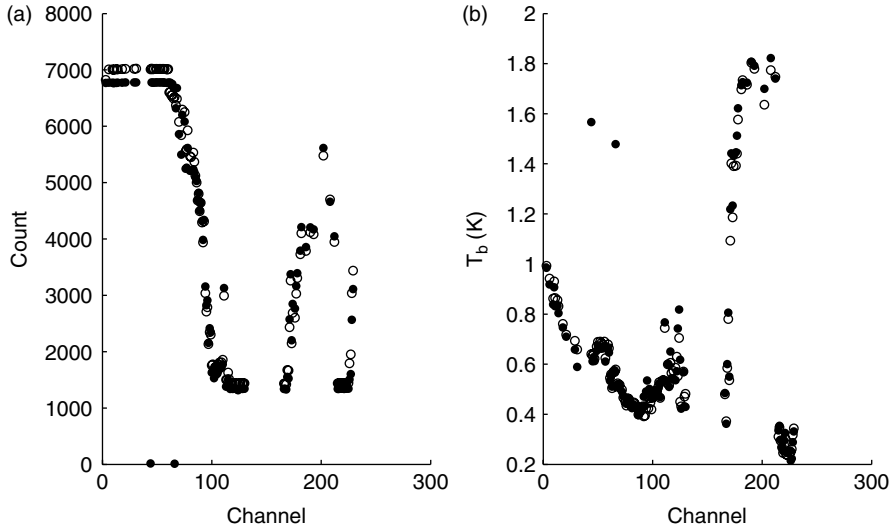

Figure 3. For indicated channels of AIRS on AQUA, (a) mean numbers of observations assimilated per day and (b) standard deviations of innovations for brightness temperatures (unit K). Filled and open circles are for OSSE and GDAS values, respectively. Channel designations are those of the renumbered subset employed at NCEP, with values plotted only for channels actually employed in the DAS experiments.

statistics value is 25% greater than the GDAS value, and QuikScat winds, for which the OSSE value is nearly double the GDAS one. The reasons for these exceptions remain unclear but will be investigated and corrected in future experiments. The current discrepancy, however, is not expected to impact the other results presented here greatly.

Standard deviations of T_b innovations for AIRS on AQUA appear in Figure 3. The corresponding OSSE and GDAS values are very similar. The only exceptions occur for channels 44 and 66 (equivalent to AIRS designations 92 and 151, respectively), which are associated with very small observation counts such that their differences are statistically insignificant. Agreements for the other radiance instruments and platforms not shown are like those for AIRS, or better. The only exceptions are channels 9 and 14 on AMSU–A on AQUA, for which the OSSE values are respectively 30% larger and smaller than their GDAS counterparts.

6.4. *Horizontal correlations of innovations*

Observation innovations are generally horizontally correlated because errors in $\mathbf{x}_b$ are horizontally correlated due to atmospheric and model dynamics. For some data types, additional correlations are obtained because their instrument plus representativeness errors are also horizontally correlated. The GSI does not treat the latter correlations explicitly (it is assumed $\mathbf{R}$ is diagonal) but simply inflates the error variances assigned to observation types with assumed highly correlated observation errors so as not to draw to such types too strongly. Otherwise, the more correlated the innovations are, the greater their net effect on the analysis increments. Essentially, weakly correlated innovations in close proximity to one another will tend to cancel each others' influences on the analysis increment $\mathbf{x}_a - \mathbf{x}_b$ as their contributions are added according to the weights prescribed by the $\mathbf{K}$ matrix as in (6). Since the GSI algorithm is very effective at filtering spatially uncorrelated observation errors, if simulated observation errors are too weakly correlated horizontally compared with their real counterparts then the OSSE will tend to create less analysis error and less subsequent background error. Thus, in order for the effects of simulated observations to be similar to those of real ones, it is important that the corresponding spatial correlations be similar.

A sample of correlations of innovations for selected observation types and channels or pressure levels are presented in this subsection. The results are presented only for particular geographical regions. Values are produced by aggregating cross-products in bins of separation distances (s). The nth bin includes distances $40(n-1) < s \leq 40n$ for s measured in km. In the figures, the first two bins are excluded because values for such close observations are undersampled. For other bins, labelled distances are their centre values of s. For conventional data, observations with pressures within 2 hPa of the indicated pressures are included in the correlation determinations.

Horizontal correlations of OSSE and GDAS innovations for a sample of conventional observations appear in Figure 6. These are all evaluated for the region north of 20°N.

Results for GOES visible and IR winds at 300 hPa appear in Figure 6(a). The two curves are very similar, with more notable differences at $s < 140$ km and $s > 500$ km, although in the latter range the values are less than 0.1 for both results. Similar agreement is obtained for all other levels in the range $p < 850$ hPa. It was necessary to correlate the added observation errors in the OSSE to obtain such agreement, otherwise the OSSE innovation correlations were only small fractions of the GDAS results. An exception to this good agreement appears in the result for 850 hPa shown in Figure 6(b), where correlations for the OSSE are about half those for the GDAS at the same distance. This discrepancy reveals a failure of the automated tuning for the horizontal observation-error correlation functions at this level. Results for the northward wind component (v) are similar to those for u at corresponding levels.

Representative results for rawinsondes also appear in Figure 6. Correlations at short distances are less than for GOES winds, presumably because the observational errors themselves have little horizontal correlation and are simulated with none. For T_v at 700 hPa, shown in Figure 6(c), the remarkable agreement, along with the good agreement of the corresponding innovation standard deviations, suggests that the background-error correlations

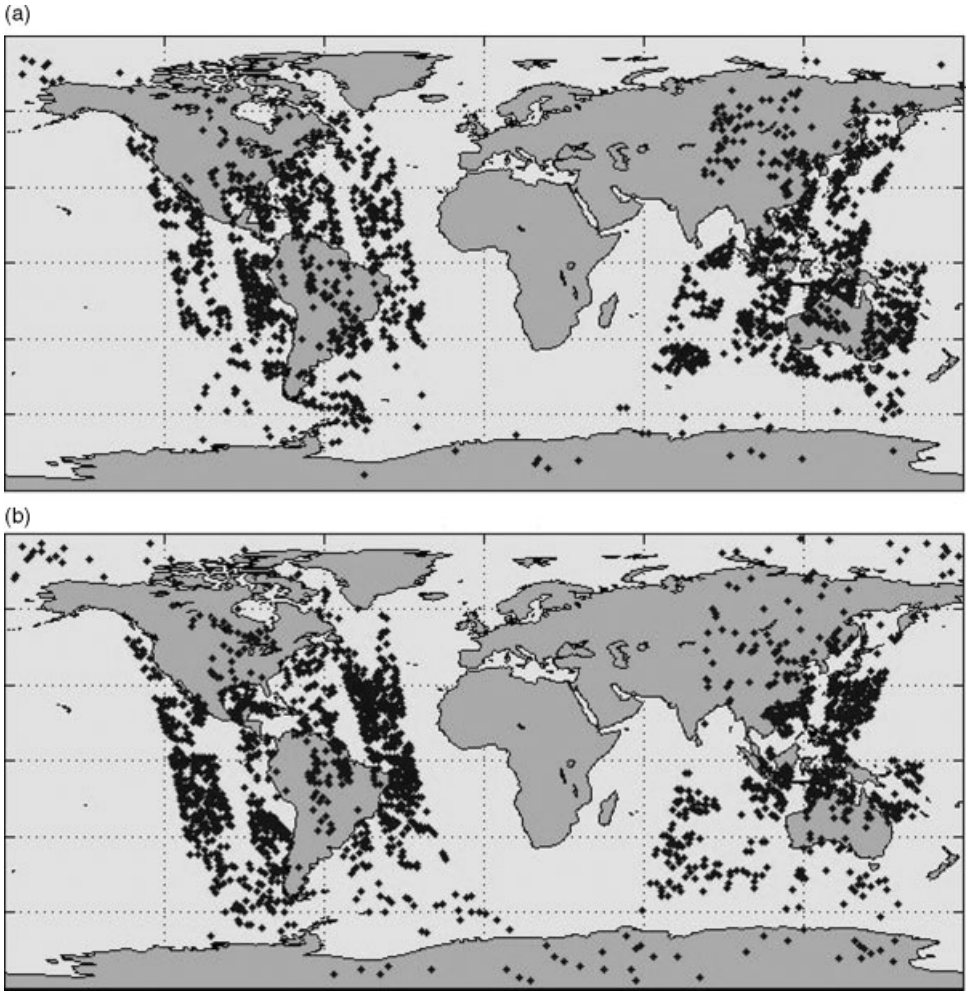

Figure 4. Locations of observations assimilated for the 6 h period centered on 1800 UTC on 12 July 2005 for channel 295 for the AIRS instrument on the AQUA satellite for (top) the OSSE and (bottom) the GDAS.

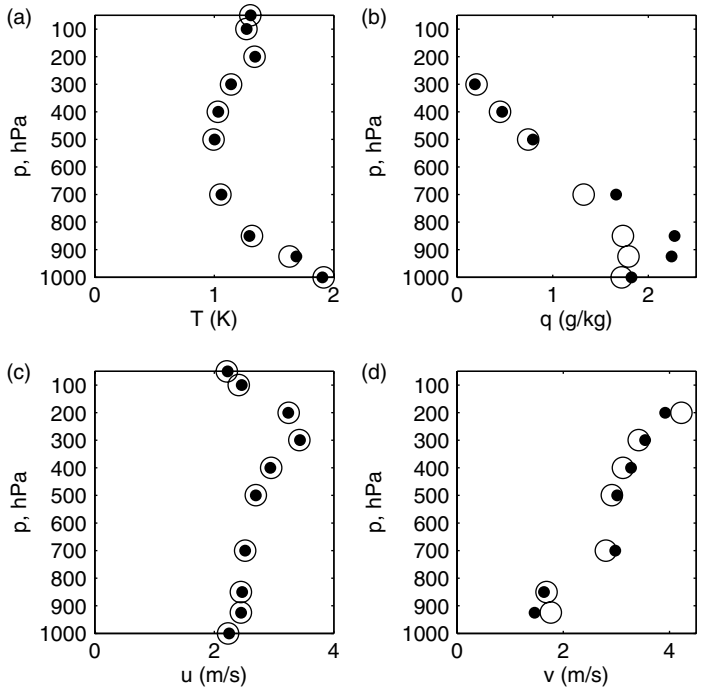

Figure 5. Standard deviations of innovations for (a) T_v, (b) q and (c) u for rawinsondes and (d) u for GOES winds north of 20°N as functions of pressure for the OSSE (filled circles) and GDAS (open circles). Units are K, g kg^{-1}, m s^{-1} and m s^{-1}, respectively.

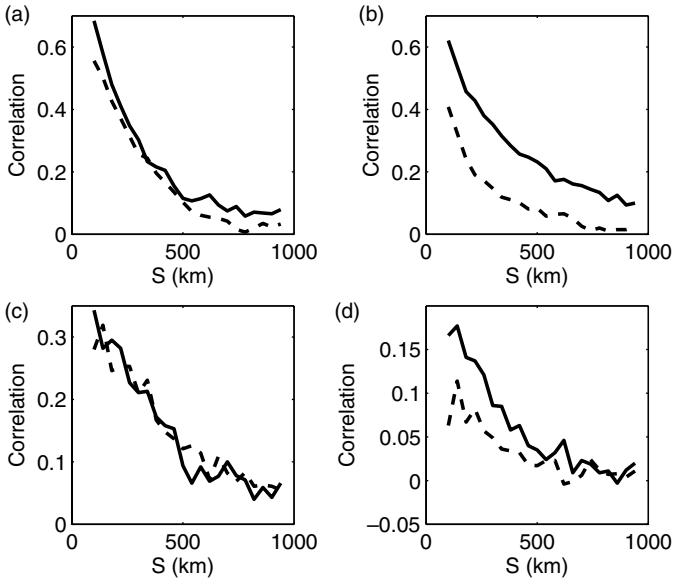

Figure 6. Horizontal correlations of innovations for OSSE (dashed lines) and GDAS (black lines) for GOES u winds near (a) 300 hPa and (b) 850 hPa and rawinsonde (c) T_v at 700 hPa and (d) u at 850 hPa. All are evaluated for the globe north of 20°N.

at this level are similar in the OSSE and GDAS since what is shown here is primarily determined by those correlations. As representative of wind throughout the troposphere, u at 850 hPa is presented in Figure 6(d). For it, OSSE values are about 70% of corresponding GDAS ones, although all values are quite small. This discrepancy suggests that the background-error correlations for wind may be weaker in the OSSE than in the GDAS. These comments about background-error correlations for T_v and wind are directly suggested only for those well-observed regions where dense rawinsonde observations exist.

As representative of horizontal innovation correlations for radiance data, results for channels 5, 6 and 8 of AMSU–A on NOAA–15 and channel 5 of HIRS–3 on NOAA–16 appear in Figure 7. The results are typical to those for other channels and instruments. The added OSSE observation errors for all these included a small component that was horizontally correlated.

6.5. *Statistics of analysis increments*

A major short-term goal guiding this work was to produce better agreement between temporal standard deviations of corresponding OSSE and GDAS analysis increments than had been obtained for the previous NCEP OSSE (Errico *et al.*, 2007a). For that work, OSSE values for the northern hemisphere were typically 30% larger than corresponding values for the real DAS and corresponding values in the southern hemisphere were 30% smaller, such that it appeared that the hemispheres were interchanged in the OSSE compared with the GDAS results!

In this subsection, all statistics are computed for only the aggregate of 0000 and 1200 UTC analysis. These two times differ from 0600 and 1800 UTC because the former pair includes rawinsondes while the latter has very few. By averaging over all times, effects of the rawinsondes will thereby be diminished and appear less apparent.

The square roots of zonal means of the temporal variances of analysis increments of T appear in Figure 8(a) and (b) for both OSSE and GDAS. Note that structures of the two sets of statistics are similar, with local maxima appearing close to the same levels. The magnitudes for the OSSE, however, are typically 5–10% smaller than GDAS ones at the same locations.

The results of the same calculation, but applied to v, appear in Figure 8(c) and (d). Here, differences are more in the 10–20% range. The patterns are also more notably different than for T. Near the South Pole, the OSSE has values only slightly larger elsewhere over Antarctica, unlike GDAS which has a global maximum. Similar comments apply for the u-field statistics (not shown).

The corresponding results for q appear in Figure 8(e) and (f). Note that, as for T, the OSSE and GDAS structures are very similar. Also, the differences in magnitude are again 5–10%, except near 900 hPa near the Equator, where the local minimum in the OSSE is approximately 20% of its GDAS value.

As examples of zonal variability of the temporal standard deviations of the analysis increments, OSSE and GDAS values are presented for T near 850 hPa and u near 500 hPa in Figure 9. For T, local maxima appear at isolated rawinsonde stations as well as in regions where the rawinsonde and aircraft observation network is dense. The local maxima do not appear at identical locations in the OSSE and GDAS, although there are many common locations such as at the rawinsonde stations along the coast of Antarctica. For the u field shown, local maxima also appear over many oceanic regions not associated with either aircraft or rawinsonde reports, although there are still many corresponding structures in the GDAS and OSSE results. The same is true for both u and v throughout most of the troposphere and for T above 500 hPa. Values for OSSE results are typically smaller than GDAS ones, in agreement with Figure 8.

Agreement between means of analysis increments in OSSE and GDAS contexts should not be expected. Most sources of bias have been omitted from the simulated observation errors. Also, the analysis-error bias introduced due to differences between the climatologies of the NR and GEOS–5 models need not be similar to the bias introduced in the GDAS due to the difference between climatologies of the real atmosphere and GEOS–5 model, unless the deficiencies in

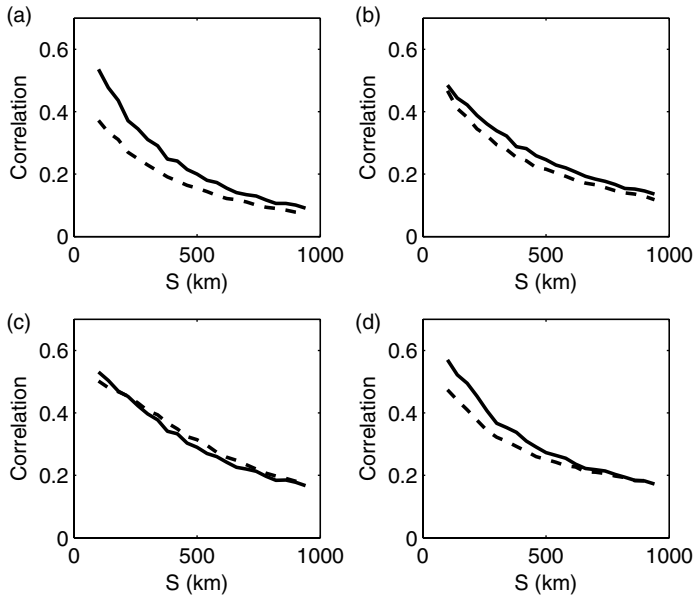

Figure 7. Horizontal correlations of innovations for OSSE (dashed lines) and GDAS (black lines) for channels (a) 5, (b) 6 and (c) 8 of AMSU–A on NOAA–15 and (d) channel 5 of HIRS–3 on NOAA–16.

the NR model climatology are negligible by comparison. For this statistic, therefore, qualitative agreement is sufficient, i.e. typical magnitudes and perhaps horizontal scales.

The pairs of fields shown in Figure 10 are typical of the agreements between temporal mean analysis increments in the OSSE and GDAS. For the T fields near 850 hPa, characteristic magnitudes and shapes of structures in the fields appear very similar in the OSSE and GDAS. More remarkable is the result that the two difference fields appear highly correlated: many relative maxima and minima appear in nearly the same locations. A similar degree of correlation is observed for T at other tropospheric levels. Correlation, to only a slightly lesser degree, is observed for the mean u increments near 200 hPa also shown in Figure 10. These correlations therefore suggest that the climatologies of the real atmosphere and NR may be more like each other than either is like that of GEOS–5.

6.6. Other results

In earlier experiments at the GMAO that had been performed without applying any horizontal correlation for observation errors, it was noticed that although the variances of uncorrelated errors added to the simulated observations could be tuned so that innovation variances obtained in the OSSE matched those in the GDAS, variances of analysis increments were always much weaker in the OSSE. In a search for what missing characteristic of the simulated errors could explain why one statistic but not the other matched, spatial correlations of the observation innovations were investigated. This revealed that while such correlations for rawinsondes matched reasonably well, those for cloud-tracked winds in particular, but also for AMSU–A (and correlations between channels for AIRS), were all much weaker in the OSSE. This weakness directly resulted in diminished magnitudes of analysis increments, especially over ocean areas where rawinsondes are absent.

The first validation studies performed for this present GMAO OSSE system were conducted for the NR simulated month of January. As for the July case, it was fairly easy to match observation innovation standard deviations somewhat. Horizontal correlations of observation errors for SATWND and AMSU–A and channel correlations for AIRS were also required to match GDAS observation-increment correlations and to increase variances of analysis increments better to match the corresponding GDAS values. None of the matches, however, were as good as shown here for July. In particular, the match of standard deviations of analysis increments for January was fairly poor for T in midlatitudes at midtropospheric levels. Validation statistics based on forecast-skill scores, to be presented in a separate report, revealed even less agreement. All these January results suggested that both background and analysis-error variances for the January simulation were significantly under-represented, potentially invalidating the OSSE for some intended applications. In an effort to understand those initially disappointing results, the present July case was examined.

7. Summary

Algorithms, software and datasets for generating a baseline set of simulated observations for future OSSE experiments have been produced at the GMAO. These are based on the nature run provided by ECMWF for the 'Joint OSSE' project. They have been validated in the context of the GMAO version of the NCEP/GMAO GSI DAS for July 2005. Metrics discussed in this report include some statistics of observation innovations and analysis increments within the troposphere.

A simple procedure was developed for efficiently tuning parameters that determine the simulated errors added to the simulated observations so that some basic statistics of observation innovations obtained in the OSSE closely match the corresponding statistics in an assimilation of real observations. In particular, for those observation types on which the tuning was focused, the OSSE innovation standard deviations were successfully induced to match within a few per cent of their GDAS values. For all radiance, SATWND and conventional surface observations,

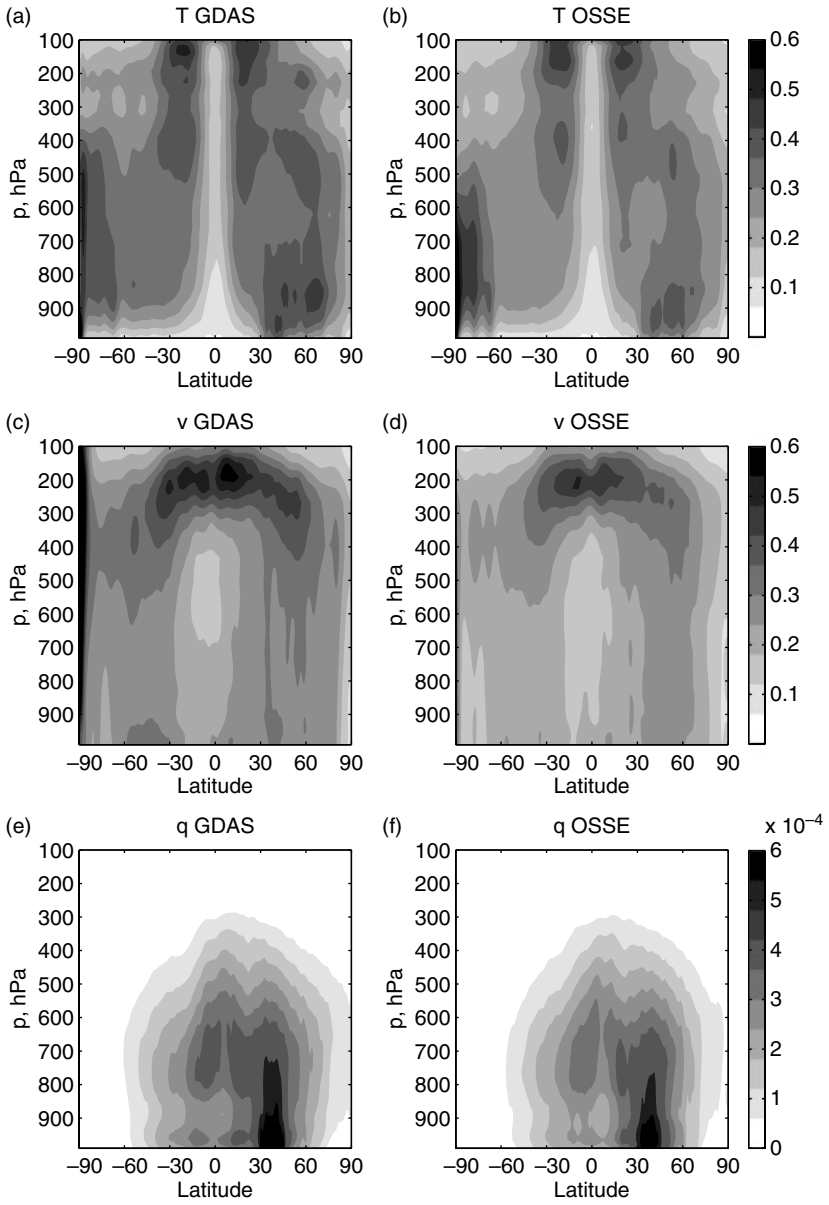

Figure 8. Square roots of zonal means of temporal variances of analysis increments of indicated fields and experiments. Units of K for T, m s^{-1} for v and g kg^{-1} for q.

the standard deviations of these added errors were less than the corresponding values specified for the observation-error statistics in the DAS. For AIRS and SATWND observations, the ratios of the corresponding deviations were typically half or less. These tuning parameters are consistent with the fact that (1) a portion of representativeness error is already implicitly included in the simulated observations because they were computed from data on a different grid than employed in the DAS and (2) the standard deviations specified as the DAS observational error statistics are inflated for observations having errors strongly correlated either horizontally or between channels.

In order to match the horizontal correlations of observation innovations for some observation types, it was necessary to create a horizontally correlated component for a portion of the simulated added observation error. This was also necessary to improve the agreement between corresponding analysis-increment statistics in the GDAS and OSSE, since innovation correlations profoundly affect the error-filtering ability of the DAS algorithm. The observation type with the largest fraction of horizontally correlated error component (larger than 0.5 at most pressure levels) was SATWND. A smaller fraction (typically 0.25) was required for AMSU–A observations. No such correlated error was added to other conventional observations. Simulated added observation errors for AIRS were correlated between channels for analogous reasons.

The result that parameters in the model simulating added observation errors can be effectively tuned so that some statistics of observation innovations can appear realistic is useful, but alone is not strong evidence of OSSE validity. First, the matching reported here is incestuous, in that the same dataset is used to derive the error model parameters as is used to validate the result. So, for example, if the OSSE yields unrealistically weak contributions to the observation innovation covariances by the background errors for some reason, the procedure will successfully inflate the observation-error component of the innovation covariances as compensation. This is precisely what was observed in the earlier GMAO OSSE validation studies

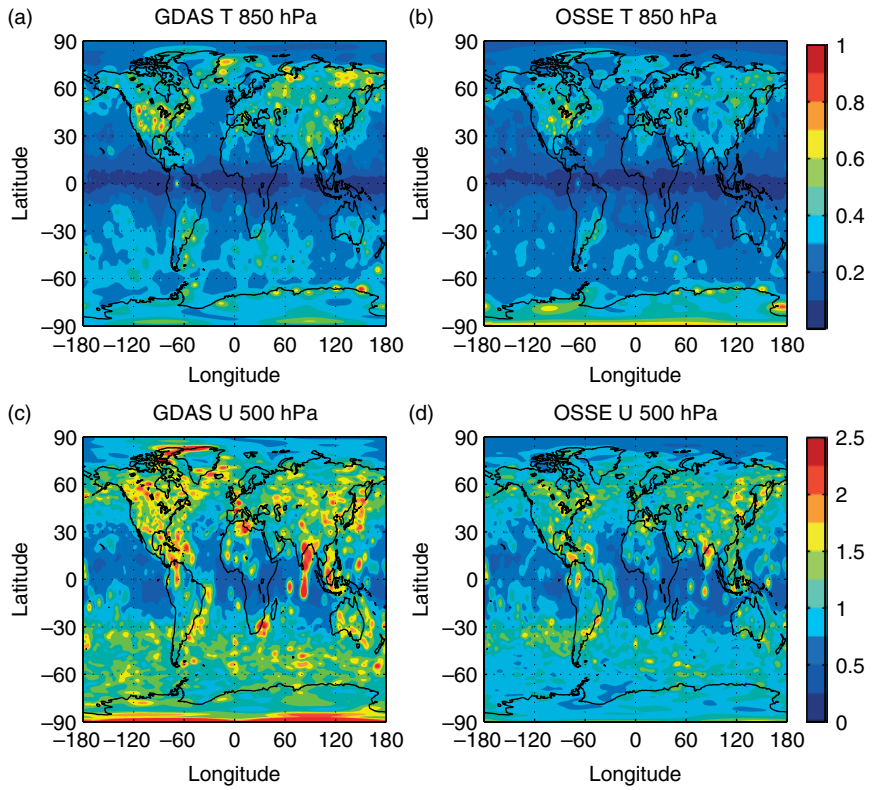

Figure 9. Temporal standard deviations of analysis increments of indicated fields, experiments and η levels. Units of K for T and m s^{-1} for u.

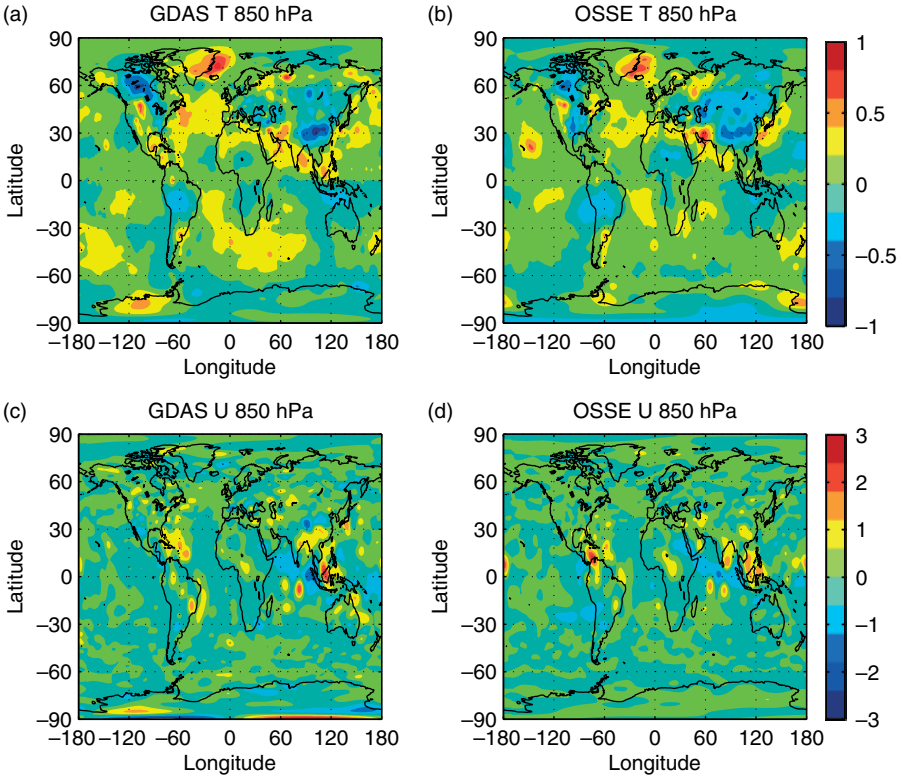

Figure 10. As in Figure 9 except for temporal means.

conducted for a January period. The covariances may be matched, but not for the correct reasons. It is therefore imperative to extend the validation to other statistics.

As a set of metrics describing the net effect of all observations within the assimilation cycles, standard deviations of analysis increments were examined. The OSSE values were generally slightly weaker for most fields in most regions compared with their GDAS counterparts. Patterns of the fields of standard deviations were similar otherwise, although local maxima did not actually coincide except at some isolated rawinsonde stations. It is unclear what characteristics of the observation innovations that

determine the analysis increments remain poorly simulated, since many statistics of the former are generally well-matched. Insufficient model error, which in this context means weaker differences in the formulations of the GEOS–5 and NR models compared with the differences between the GEOS–5 model and real atmospheric dynamics and physics, can explain the result.

Temporal mean values of analysis increments were also examined. Corresponding values for many fields had similar magnitudes and appeared highly correlated to each other. This similarity of means also suggests that the NR climatology that helps determine such means is more similar to the atmosphere than the GDAS–5 climatology is. Since these mean fields are determined in large part by systematic errors in the DAS, particularly by forecast modelling errors, this suggests that the GEOS–5 model error with respect to atmospheric behaviour may be significantly greater than the error in the NR model. While this cannot be concluded, it must be hypothesized.

The results for this new baseline OSSE system are certainly in better general agreement with results from assimilating real observations than have been reported for past OSSE validations. For some OSSE applications, especially those concerned with qualitative assessments, the present baseline may be quite adequate. It is also an appropriate initial baseline from which to measure improvements in new baseline datasets to be produced. The largest discrepancies of statistics in this study concern observations that have little demonstrated impact on the analysis, such as for QUIKSCAT and SSM/I ocean-surface winds. The present results therefore suggest the OSSE's utility but also the difficulties in producing an OSSE that validates adequately.

Given the importance of spatial and channel correlations of observation error, several improvements should be made to the present error-simulation model. Other spectral shape functions defining the horizontal correlations should be considered. Also a more appropriate but still computationally efficient way to introduce horizontal and vertical (or channel) correlations simultaneously should be developed, with a re-tuning of all correlation model parameters.

Simulations of observations and added error for some observation types also need improvement. Although QUIKSCAT will be dropped from future baseline sets of observations, the problems with it suggest that further effort is required for any similar observations. Although for this study the locations of SATWND observations do not necessarily coincide with trackable features in the nature run, this neglect will be corrected in the near future. Also, the OSSE will be extended to include most present observation types.

Future reports concerning the development of baseline sets of observations for the GMAO OSSE will include additional metrics not discussed here. Impacts of observation subsets on forecast skill will primarily be compared (Privé *et al.*, 2012) by using estimates provided by adjoints of the assimilation and forecast system (Langland and Baker, 2004; Gelaro and Zhu, 2009). Although this latter approach has additional limitations not present in the data denial one (Gelaro and Zhu, 2009), it can present a more detailed picture than is computationally feasible otherwise. Also, a study of analysis error in the OSSE context is currently under way, made possible by the fact that in such experiments the truth is given precisely by the nature run.

Acknowledgements

The authors benefited from many conversations with Ronald Gelaro, Will McCarty and Arlindo da Silva. Joanna Joiner suggested using an elevated surface to provide a cloud effect for IR radiances. Steve Bloom and Tong Zhu provided some useful software. Ravi Govindaraju, Joesph Stassi and Thomas Owens provided some additional necessary datasets. The ECMWF nature run was provided by Erik Andersson through arrangements made by Michiko Masutani. Support for this project was encouraged by Michele Rienecker and provided by grant numbers WBS 802678.02.17.01.06 and MAP/04–0000–0080 from the National Aeronautics and Space Administration and grant number ATM–0745906 from the National Science Foundation.

Appendix A. The radiance-affecting cloud-presence model

For several reasons stated in the main text, the algorithm indicating the presence of radiance-affecting clouds is kept simple here. The only cloud information from the NR used is the cloud fractional coverage f in the NR grid boxes specified for high, mid and low levels. At each observation location, the algorithm considers these levels in succession, stopping if any cloud effect is obtained.

For each of the three levels j, the probability P_j that a cloud is affecting some radiance channels is defined using a simple, piecewise linear function of f_j:

$$P_j = \begin{cases} 0. & \text{if } f_j \le a, \\ 0.5(f_j - a_j)/(b_j - a_j) & \text{if } a_j < f_j \le b_j, \\ 0.5 + 0.5(f_j - b_j)/(c_j - b_j) & \text{if } b_j < f_j < c_j, \\ 1. & \text{if } f_j \ge c_j, \end{cases}$$

where a_j, b_j, c_j are parameters to be tuned to yield the desired QC-accepted observation counts. Corresponding to each determined value of P_j, a random number r_j is drawn from a uniform distribution on the interval $0 \le r_j \le 1$. An affecting cloud is declared present if $r_j < P_j$. In that case, the cloud-top pressure defining the effective radiative surface is assigned to be $p_c = \sigma_j p_s$, where p_s is the true surface pressure at that location and σ_j is another tuning parameter.

For MW instruments, the same procedure is used except that the three cloud fractions are replaced by the convective and stratiform precipitation rates at the surface. The values of precipitation rates and cloud fractions used in these calculations are actually those spatially and temporally interpolated to observation locations, the same as for other NR fields. Values of a, b, c are identically set to 0.001 (metres of water per m^2 area per three-hour period) with σ set to 0.7 and 0.85 for convective and stratiform precipitation, respectively. Values of the tunable parameters for HIRS and AIRS appear in Table A1.

Appendix B. The procedure for tuning explicit observation error parameters

The variances of added errors for the OSSE are tuned so that innovations of the same type and pressure level or channel have similar variances $\mathbf{V}$ in both the GDAS and OSSE. In the GDAS,

$$\mathbf{V}_t = \mathbf{R}_t + \mathrm{diag}\left(\mathbf{H}\mathbf{B}_t\mathbf{H}^T\right) , \qquad \text{(B1)}$$

where $\mathbf{H}$ and $\mathbf{H}^T$ are the tangent linear and adjoint versions of H, $\mathbf{B}_t$ is the true background-error covariance and $\mathbf{R}_t$ is

Table A1. Parameters used for determining the presence of clouds affecting simulated IR radiance observations for the OSSE.

Instrument	j	Level	a	b	c	σ
AIRS	1	high	0.10	0.40	0.70	0.35
AIRS	2	mid	0.15	0.25	0.35	0.65
AIRS	3	low	0.20	0.25	0.30	0.85
HIRS	1	high	0.10	0.40	0.70	0.35
HIRS	2	mid	0.15	0.15	0.15	0.65
HIRS	3	low	0.20	0.20	0.20	0.85

the true observation-error variance. These may differ from the corresponding covariances specified in the DAS. In fact, for SATWND and hyper-spectral instruments such as AIRS, the innovation variances are less than a quarter of the value of corresponding elements of **R**, because the latter have been inflated to mitigate negative effects in the DAS due to its neglect of observation-error correlations.

In contrast, in the OSSE,

$$\mathbf{V}_\mathrm{o} = \mathbf{R}_\mathrm{e} + \mathbf{R}_\mathrm{i} + \mathrm{diag}\left(\mathbf{H}\mathbf{B}_\mathrm{o}\mathbf{H}^\mathrm{T}\right) , \tag{B2}$$

where $\mathbf{R}_\mathrm{e}$ and $\mathbf{R}_\mathrm{i}$ are the variances of observation errors added explicitly and representativeness errors introduced implicitly, respectively, and $\mathbf{B}_\mathrm{o}$ is the actual (rather than specified) background-error covariance realized in the OSSE. The $\mathbf{R}_\mathrm{i}$ values arise because the NR and DAS grids and the CRTM versions used to create and assimilate observations differ.

The background errors obtained in the DAS are determined by forecast model error and propagated analysis errors. The latter are partly determined by observation errors. Therefore, $\mathbf{B}_\mathrm{o}$ in (B2) should be an implicit function of the observation-error statistics. Without knowing this function, it not obvious how to adjust trial values of $\mathbf{R}_\mathrm{e}$ to obtain a desired value of $\mathbf{V}_\mathrm{o}$.

Fortunately, it was observed in early tuning experiments that the portion of $\mathbf{V}_\mathrm{o}$ due to background errors did not change much as $\mathbf{R}_\mathrm{e}$ was reasonably varied. By assuming that $\mathbf{R}_\mathrm{i}$ and $\mathbf{B}_\mathrm{o}$ are unchanged, a simple tuning algorithm was used very successfully. An updated estimate of $\mathbf{R}_\mathrm{e}$ is determined from a previous experiment that yielded innovation variances $\tilde{\mathbf{V}}_\mathrm{o}$ produced using added errors having variances $\tilde{\mathbf{R}}_\mathrm{e}$:

$$\mathbf{R}_\mathrm{e} = \mathbf{V}_\mathrm{t} - \tilde{\mathbf{V}}_\mathrm{o} + \tilde{\mathbf{R}}_\mathrm{e}. \tag{B3}$$

In our experience, at most two iterations of this procedure were required to obtain $\mathbf{V}_\mathrm{t} \approx \mathbf{V}_\mathrm{o}$ as presented in this report.

Additional parameters required to create horizontally correlated added observation errors were specified by fitting Gaussian-shaped functions to computed horizontal correlations of innovations for each observation instrument and channel or level as functions of horizontal separation s (i.e. assuming isotropy). The fitting is weighted in s to emphasize better fits for medium distances that are expected to have greater impact on DAS results (because not many observations of the same type are very close together and combinations of widely separated observations receive little weight in the DAS). The L_h for the added OSSE errors are specified as the corresponding values obtained from fitting GDAS results. The fitting also provides an extrapolation of the correlation to separation distance $s = 0$ (as in Hollingsworth and Lonnberg, 1986) that yields a fraction γ of variance for the correlated part of **d**, including both background and observation-error components, to the total variance. The ratios of variances for correlated versus total error to be added are then estimated as

$$\nu_\mathrm{e} = \frac{\mathbf{V}_\mathrm{t}\gamma_\mathrm{t} - \tilde{\mathbf{V}}_\mathrm{o} + \tilde{\mathbf{R}}_\mathrm{e}}{\mathbf{R}_\mathrm{e}} , \tag{B4}$$

where the subscript t indicates target values provided by the GDAS results. The quantity in the numerator in (B4) is simply an estimate of the desired variance for the correlated part of the observation error to be added based on the GDAS and previous OSSE results.

Appendix C. Power spectra of the horizontally correlated random fields determining observational errors

The horizontal correlation model for added observational errors is defined as

$$r = \exp\left(-0.5s^2/L_\mathrm{h}^2\right) , \tag{C1}$$

where s is the great-circle separation distance on the sphere and L_h is the desired horizontal length-scale. The error statistics are assumed to be geographically independent. As shown in Weaver and Courtier (2001), the expected power spectrum, triangularly truncated at wave number N, for random fields on the sphere approximately having such correlations is

$$P_n = \beta(2n+1)\exp\left[-0.5a^{-2}L_\mathrm{h}^2(n^2+n)\right] , \tag{C2}$$

where P_n is the power for spherical harmonics of order $n = 0, \ldots, N$, a is the Earth's radius and β is the factor that yields the desired global-mean variance V:

$$\beta = V\left(\sum_{n=0}^{N} P_n\right)^{-1} . \tag{C3}$$

Random spherical harmonic coefficients c_n^m of zonal wave number $m = 0, \ldots, N$ and $n = m, \ldots, N$ are determined consistently with this expected power spectrum by setting $c_n^m = c_\mathrm{r} + \mathrm{i}c_\mathrm{i}$, where $\mathrm{i} = \sqrt{-1}$ and c_r and c_i are random variables drawn separately for each m and n from distributions with 0 mean and variances $(4n+1)P_n$. The factor before P_n here is the number of real or imaginary components (excluding the imaginary ones defined for $m = 0$ that are constrained to be zero-valued) for a given n, including those implied for $m = -N, \ldots, -1$. The latter inclusion is necessary because coefficients $c_n^m = c_n^{-m*}$, where the asterisk indicates a complex conjugate, are implied to constrain the random fields to be real-valued.

References

Bacmeister JT, Suarez MJ, Robertson FR. 2006. Rain re-evaporation, boundary-layer/convection interactions and Pacific rainfall patterns in an AGCM. *J. Atmos. Sci.* **63**: 3383–3403.

Bloom S, Takacs L, DaSilva A, Ledvina D. 1996. Data assimilation using incremental analysis updates. *Mon. Weather Rev.* **124**: 1256–1271.

Bormann N, Andrew C, Bauer P. 2010. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II: Application to AIRS and IASI data. *Q. J. R. Meteorol. Soc.* **136B**: 1051–1063.

Daley R, Menard R. 1993. Spectral characteristics of Kalman filter systems for atmospheric data assimilation. *Mon. Weather Rev.* **121**: 1554–1565.

Dee DP. 2005. Bias and data assimilation. *Q. J. R. Meteorol. Soc.* **120**: 3332–3343.

Errico RM, Yang R, Masutani M, Woollen J. 2007a. Estimation of some characteristics of analysis error inferred from an observing system simulation experiment. *Meteorol. Z.* **16**: 695–708.

Errico RM, Bauer P, Mahfouf JF. 2007b. Issues regarding the assimilation of cloud and precipitation data. *J. Atmos. Sci.* **64**: 3785–3798.

Gelaro R, Zhu Y. 2009. Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus* **61A**: 179–193.

Han Y, van Delst P, Liu Q, Weng F, Yan B, Treadon R, Derber J. 2006. '*JCSDA Community Radiative Transfer Model (CRTM) – Version 1*', NOAA Tech. Report 122. U.S. Department of Commerce: Washington, DC.

Hollingsworth A, Lonnberg P. 1986. The statistical structure of short–range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus* **38A**: 111–136.

Ide K, Courtier P, Ghil M, Lorenc AC. 1997. Unified Notation for data assimilation: operational, sequential and variational. *J. Meteorol. Soc. Jpn* **75**: 181–189.

Kleist DT, Parrish DF, Derber JC, Treadon R, Wu W-S, Lord S. 2009. Introduction of the GSI into the NCEP global data assimilation system. *Weather and Forecasting* **24**: 1691–1705.

Langland RH, Baker NL. 2004. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus* **56A**: 189–201.

Lin S-J. 2004. A vertically Lagrangian finite-volume dynamical core for global models. *Mon. Weather Rev.* **132**: 2293–2307.

Masutani M, Riishojgaard L-P, Schlatter TW, Woollen JS, Terry J, Toth Z, Errico R, Emmitt GD, Reale O, Andersson E, Stoffelen A, Marseille G-J, Xie Y, Song Y, Liu E, Jusem J-C, Greco S, Wood SA, Yang R, McConaughy G, Devenyi D, Privé N, Weygandt SS, Sun H, Lord S. 2007. 'Progress in Observing Systems Simulation Experiments (a new nature run and international collaboration)'. In *Proceedings of the 22nd Conference on Weather Analysis and Forecasting and the 18th Conference on Numerical Weather Prediction, Park City, Utah*, 25–29 June 2007. American Meteorological Society: Boston, MA.

McCarty W, Errico RM, Gelaro R. 2012. Cloud coverage in the joint OSSE nature run. *Mon. Weather Rev.* **140**: 1863–1871.

Privé NC, Errico RM, Tai K-S. 2012. Validation of forecast skill of the Global Modeling and Assimilation Office Observing System Simulation Experiment. *Q. J. R. Meteorol. Soc.* DOI: 10.1002/qj.2029.

Reale O, Terry J, Masutani M, Andersson E, Riishojgaard L-P, Jusem J-C. 2007. Preliminary evaluation of the European Centre for Medium-range Weather Forecasts (ECMWF) nature run over the tropical Atlantic and African monsoon region. *Geophys. Res. Lett.* **34**: L22810. DOI:10.1029/2007GL031640.

Rienecker MM, Suarez MJ, Todling R, Bacmeister J, Takacs L, Liu H-C, Gu W, Sienkiewicz M, Koster RD, Gelaro R, Stajner I, Nielsen JE. 2008. '*The GEOS–5 Data Assimilation System – Documentation of versions 5.0.1, 5.1.0, and 5.2.0*', Technical Report Series on Global Modeling and Assimilation No. 27, Suarez M, ed., NASA/TM–2008–104606, 109 pp. NASA: Washington, DC.

Weaver A, Courtier P. 2001. Correlation modeling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.* **127**: 1815–1846.

Wu W-S, Purser RJ, Parrish DF. 2002. Three dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Weather Rev.* **130**: 2905–2916.

